

DOCUMENT RESUME

ED 281 852

TM 870 223

AUTHOR Heausler, Nancy L.
TITLE A Primer on MANOVA Omnibus and Post Hoc Tests.
PUB DATE 30 Jan 87
NOTE 21p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (Dallas, TX, January 30, 1987).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Analysis of Variance; Discriminant Analysis; Factor Analysis; *Hypothesis Testing; *Multivariate Analysis; *Statistical Significance; Statistical Studies
IDENTIFIERS Omnibus Test; Post Hoc Methods

ABSTRACT

Each of the four classic multivariate analysis of variance (MANOVA) tests of statistical significance may lead a researcher to different decisions as to whether a null hypothesis should be rejected: (1) Wilks' lambda; (2) Lawley-Hotelling trace criterion; (3) Roy's greatest characteristic root criterion; and (4) Pillai's trace criterion. These four omnibus test statistics are discussed and their optimal uses illustrated using hypothetical data sets. Discriminant analysis as a post hoc method to MANOVA is illustrated in detail. Once a significant MANOVA has been found, the next step is to interpret the non-chance association between dependent and independent variables. (Author/GDC)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

A Primer on MANOVA Omnibus and Post Hoc Tests

Nancy L. Heausler

University of New Orleans

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☒ This document has been reproduced as
received from the person or organization
originating it.

☐ Minor changes have been made to improve
reproduction quality.

• Points of view or opinions stated in this docu-
ment do not necessarily represent official
OERI position or policy.

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

N. L. Heausler

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

Paper presented at the annual meeting of the Southwest
Educational Research Association, Dallas, January 30,
1987.

BEST COPY AVAILABLE

Abstract

Each of the four classic MANOVA test criteria may lead a researcher to different decisions as to whether a null hypothesis should be rejected. These various omnibus test statistics are discussed and their optimal uses are illustrated using hypothetical data sets. Once a significant MANOVA has been found, the next step is to interpret the non-chance association between dependent and independent variables. Discriminant analysis as a post hoc method to MANOVA is illustrated in detail.

Analysis of variance (ANOVA) is a common statistical method. However, a problem with analysis of variance and other univariate methods is that only one dependent variable can be considered in a given analysis. Since most researchers usually care about multiple effects, and since most effects are multiply caused, multivariate methods are often vital in behavioral research (Thompson, 1986, p. 13). With multivariate analysis of variance (MANOVA), more than one dependent variable can be considered simultaneously. The full network of variable relationships are considered in MANOVA. This ususally would more accurately reflect the reality about which researchers attempt to generalize (Thompson, 1986).

The mathematics underlying MANOVA are complex and handled by matrix algebra. But with the advent of the computer age, multivariate techniques became available for practical application by more researchers (Barker & Barker, 1984). MANOVA differs from ANOVA in that the MANOVA statistical test of significance may take one of several possible forms. The purpose of the present paper is to discuss the tests of statistical significance associated with MANOVA and the appropriate uses of each. Also, the use of discriminant analysis as a post hoc method for interpreting results is discussed.

To illustrate the discussions, two hypothetical

data sets are presented. Table 1 presents the data for the first hypothetical research situation. TRE and GRA are the independent variables. TRE refers to treatment condition with 1 denoting membership in the experimental condition and 2 denoting membership in the control group. GRA refers to grade level, respectively, first, second, and third. The dependent variables are DEPA, which refers to an achievement test score, and DEPB, which refers to a creativity test score. Notice that this is the simplest case of a two-way MANOVA, with only two dependent variables. The data were designed so that only one of the two ways has an effect, and it effects both of the dependent variables.

Insert Table 1 about here.

The second hypothetical research situation involves the same independent variables but with different data on the dependent variables. DEPC refers to another achievement test score and DEPD refers to another creativity test score. In this hypothetical case, both ways affect only one of the dependent variables.

Insert Table 2 about here.

Tests of Statistical Significance

To understand the differences in the tests of statistical significance, one must understand some of the conceptual theory underlying the MANOVA process. Comparing MANOVA with ANOVA may shed light on the multivariate technique.

At the heart of MANOVA is factor analysis (principal components). Factor analysis is used to determine the dimensionality and nature of the sets of dependent variables that are sensitive to treatment effects. Thus the ANOVA techniques and interpretations are possible for separate, independent dimensions (Barker & Barker, 1984). Thus an understanding of the role of factor analysis is necessary to understand the differences among the test criteria and the appropriate selection of one test for a given research situation.

In ANOVA, the total sum of squares is first determined. This is a measure of the extent of variability in the data set. In MANOVA, because two or more dependent variables are considered, the cross-products between the dependent variables as well as the sum of squares are taken into account. Thus, the sum-of-squares-cross-products (SSCP) matrix is calculated. The SSCP matrix is analogous to ANOVA's sum of squares.

In ANOVA, the total sum of squares is broken into two parts, the sum of squares treatment and the sum of

squares error. These two quantities are used in determining the F ratio test statistic in the fixed effects case. In MANOVA, the SSCP matrix also consists of a total that is analyzed into two parts, SSCP treatment and SSCP error. Division is performed on these matrices by way of matrix algebra, similar to the division of the F ratio. A new matrix results, and this matrix reflects the extent to which the variables discriminate between the treatment groups (Barker & Barker, 1984).

Factor analysis is applied to this new matrix. Factor scores are created, which are synthetic variables. The synthetic variables are uncorrelated and maximally separate the treatment group means. Eigenvalues are important in factor analysis. An eigenvalue represents the portion of variance of all the variables that can be accounted for by a given set of factor scores. The number of possible eigenvalues for a data set is a function of the number of dependent variables in a given research situation. The tests of statistical significance are affected differentially by the eigenvalues that result for a given data set (Barker & Barker, 1984).

The four classic MANOVA tests of statistical significance are Wilks' lambda, the Lawley-Hotelling trace criterion, Roy's greatest characteristic root criterion, and Pillai's trace criterion. To determine

Wilks' lambda, the reciprocals of one plus each eigenvalue are multiplied in succession. The Lawley-Hotelling trace criterion simply involves summing all eigenvalues. Roy's greatest characteristic root criterion applies a formula to only the eigenvalue from the first dimension, or root. Pillai's trace criterion is the same as Roy's, except that the formula is applied to each of the eigenvalues (Barker & Barker, 1984). Thus, each criterion utilizes all eigenvalues except for Roy's, which only uses the first eigenvalue.

"Trace" is an important concept in discriminating the appropriateness of each test criterion. The eigenvalues sum to the "trace". "Trace" is the discriminating variance of the variables undergoing factor analysis. If trace is almost totally located in the first eigenvalue, then there is only one strong dimension useful in discriminating the experimental groups. Trace could be distributed evenly among the all eigenvalues. All four of the test criteria are differentially affected by trace. If trace is mostly located in the first root, then Roy's test is the most sensitive. If trace is equally distributed among the eigenvalues, then Wilks' test is the most sensitive. Lawley-Hotelling and Pillai's are intermediate in sensitivity to trace (Barker & Barker, 1984).

The appropriate test criteria should be chosen a priori. If only one factor is extracted from the set

of dependent variables, then all the MANOVA tests are equivalent. However, if more than one root is extracted, then the maximally sensitive or most powerful test should be chosen. Roy's test should be used when the researcher expects that one single major dimension will be extracted. Wilks' lambda should be chosen when two or more dimensions are expected and are expected to be of relatively equal importance. Lawley-Hotelling's and Pillai's test are intermediate in sensitivity when compared to the other two criteria, and should be chosen in intermediate situations (Barker & Barker, 1984).

To make the above discussion concrete, references will be made to the hypothetical research situation. The data in the first hypothetical data set were analyzed with the MANOVA procedure from SPSSX, the updated version of the Statistical Package for the Social Sciences (Nie, Hull, Jenkins, Steinbrenner & Bent, 1975). Table 3 presents the results of the four MANOVA significance tests for the effect GRA (grade level). Table 4 lists the eigenvalues associated with the two roots extracted. Two roots are the maximum that can be extracted for this research situation, since the hypothetical data involve only two dependent variables. Notice that the value of Hotelling's criterion is the sum of the two eigenvalues.

Insert Tables 3 and 4 about here.

Examining the eigenvalues also indicates that the eigenvalue for root one is much larger than the eigenvalue for root two. This means that trace is mostly located in the first root. Thus one strong dimension seems to be affecting the dependent variables. A visual examination of the data in Table 1 shows that the data do not differ with respect to treatments 1 and 2. The differences only involve grade level. So, the one dimension affecting the dependent variables could be considered grade level.

In this case, Roy's test is the most sensitive, Wilks' is the least sensitive, and Pillai's and Hotelling's are in between. However, in this hypothetical situation, all the values listed are considered statistically significant. (To determine significance of Roy's test, appropriate tables must be consulted.) The results are not always so similar. Different decisions may be made by choosing different test statistics. It is possible that the four tests will disagree as to whether the null hypothesis should be rejected at a specified alpha level (Bray & Maxwell, 1985).

The second hypothetical data set was also analyzed using SPSSX. The results for effect GRA are presented in Table 5, and the eigenvalues are presented in Table

6. Notice that the eigenvalues of roots 1 and 2 are not as different as they were in the first hypothetical situation. Trace is more evenly distributed between the eigenvalues. All the dimensions may be affecting the dependent variables. The data were designed so that both ways, treatment and grade level, would affect one of the dependent variables. In this situation, Wilks' lambda is the most sensitive test and Roy's is the least. Notice that in this case different decisions about probability are possible with the selection of different criteria.

Insert Tables 5 and 6 about here.

In the same research situation, the results for the effect TRE (treatment condition) are presented in Table 7. Notice only one eigenvalue is extracted because there are only two treatment conditions (see Table 8). The value of Hotelling's criterion is the same as the eigenvalue. The value of Roy's is the same as Hotelling's because both use the same formula, either applied to the first root or to all roots. This is evidence that all test statistics are appropriate when only one root is extracted.

Insert Tables 7 and 8 about here.

Discriminant Analysis as a Post Hoc Method

The tests of statistical significance are not the only things a thoughtful researcher should consider. The tests are a way to start the researcher on the road to meaningfully interpreting results. If an independent variable is found to be statistically significant, then there is some non-chance association between that variable and the dependent variables. Some attempt must be made to interpret the association in a meaningful way (Stevens, 1972). This is especially important if the independent variable has more than two levels. Busenbark (1985) used discriminant as a post hoc to MANOVA in her dissertation. One statistically significant way had nine levels and discriminant analysis allowed the researcher to draw sensible conclusions.

Discriminant analysis is often used for classifying people into groups. In MANOVA, the people are already in groups and the discriminant analysis is used to characterize the major differences among the groups. The differences are revealed through the discriminant functions. The researcher can interpret the functions to determine the origins of effects (Stevens, 1982).

Besides the discriminant functions, the researcher should look at the structure matrix. The structure coefficients are the correlations of each of the

dependent variables with each set of discriminating function scores. The researcher should notice the variables that are highly correlated with each function. Those variables are primarily responsible for the separation among the groups in a given function. Function coefficients could be used instead of structure coefficients for determining the relative importance of the dependent variables in relation to each discriminant function (Steven, 1972). However, structure coefficients may be more useful than function coefficients because structure coefficients are correlation coefficients. Correlation coefficients have a minimum and a maximum, unlike function coefficients.

The hypothetical research situation may make the use of discriminant analysis as a post hoc method for MANOVA more understandable. For the first MANOVA data set, the effect GRA was found to be statistically significant. However, this is not enough information to make reasonable interpretations about the association between the grade levels and the dependent variables, creativity and achievement. So discriminant analysis was performed on the data. The discriminant groups are the three grade levels.

Table 9 presents the discriminant functions. The two functions are statistically significant. Table 10 lists the structure coefficients. The two dependent

variables have similar absolute magnitudes on function I, but the structure coefficient is positive for DEPA and the coefficient is negative for DEPB. The structure coefficients for DEPA and DEPB for function II are both large and positive. So, both dependent variables are responsible for the separation along this dimension. Thus, both dependant variables contribute to the separation along both dimensions, with the dependent variables having opposite effects on function I.

Insert Tables 9 and 10 about here.

The group centroids presented in Table 11 make the interpretation meaningful. A centroid is the average discriminant score of the subjects in a given group for a given function. The centroids on function I indicate that all three groups are fairly different. Notice that group 3 has a large positive centroid. Remember that DEPB has a high positive structure coefficient on function I. This would lead the researcher to expect high scores for the people in the third grade on DEPB, the creativity test score. The group 1 centroid for function I is large and negative and the DEPA structure coefficient is also large and negative. This suggests that the children in grade 1 have high scores on DEPA, the achievement test score. The group means presented

in Table 12 confirm these expectations.

Insert Tables 11 and 12 about here.

Note that groups 1 and 3 both have large positive centroids on function II. The dependent variables are equally affected by this dimension, as shown by the structure coefficients. This reveals a similarity between grades 1 and 3, but not with grade 2. Function II seems to separate grade 2 subjects from grades 1 and 3 subjects. Refer to Table 12 to see the group means. The hypothetical research situation is simple enough that the interpretations may seem redundant. In reality, with large samples and many variables, discriminant analysis is a valuable aid in interpreting the non-chance association between the independent and dependent variables analyzed by MANOVA.

Summary

With the application of MANOVA dramatically increasing in the behavioral sciences (Bray & Maxwell, 1985), researchers should make themselves increasingly familiar with its uses and interpretations. Thus, researchers should be aware of the four MANOVA omnibus test statistics and their optimal uses. Wilks' lambda is the most widely used of the four approaches (Bray & Maxwell, 1985), but, as discussed in the present paper, all tests have different sensitivity. The test should

be chosen a priori in accordance with the research situation. The researcher must make a careful selection as different tests can lead to different decisions about probability.

If an omnibus test reveals a significant effect, then discriminant analysis is available as a method for further investigating the group differences. As MANOVA involves two or more dependent variables and often several independent variables, post hoc methods for further analyzing results are valuable for meaningful interpretations. And, meaningful interpretations are the goal of all research.

References

- Barker, H. R., and Barker, B. M. (1984). Multivariate analysis of variance (MANOVA): A practical guide to its use in scientific decision making. University, AL: University of Alabama Press.
- Bray, J. H., and Maxwell, S. E. (1985). Multivariate analysis of variance. Beverly Hills: Sage.
- Busenbark, M. M. (1985). Information use in educational decisions of administrators, teachers, and parents. Unpublished doctoral dissertation, University of New Orleans.
- Nie, H. H., Hull, C. H., Jenkins, J. G., Steinbrenner, K., & Bent, D. H. (1975). SPSS: Statistical package for the social sciences (2nd ed.). New York: McGraw-Hill.
- Stevens, J. P. (1972). Four methods of analyzing between variation for the k-group MANOVA problem. Multivariate Behavioral Research, 7, 499-522.
- Thompson, B. (November, 1986). Two reasons why multivariate methods are usually vital: An understandable reminder with concrete examples. Paper presented at the annual meeting of the Mid-South Educational Research Association, Memphis, TN.

Table 1

First Hypothetical Data Set

Case TRE GRA DEPA DEPB

1	1	1	8	3
2	1	1	8	4
3	1	1	8	5
4	1	2	3	3
5	1	2	4	4
6	1	2	5	5
7	1	3	3	9
8	1	3	4	9
9	1	3	5	9
10	2	1	8	3
11	2	1	8	4
12	2	1	8	5
13	2	2	3	3
14	2	2	4	4
15	2	2	5	5
16	2	3	3	9
17	2	3	4	9
18	2	3	5	9

Table 2

Second Hypothetical Data Set

Case TRE GRA DEPC DEPD

1	1	1	9	3
2	1	1	8	1
3	1	1	7	8
4	1	2	8	3
5	1	2	7	6
6	1	2	6	4
7	1	3	7	2
8	1	3	6	9
9	1	3	5	2
10	2	1	6	5
11	2	1	5	7
12	2	1	4	6
13	2	2	5	2
14	2	2	4	4
15	2	2	3	7
16	2	3	4	1
17	2	3	3	2
18	2	3	2	7

Table 3

Multivariate Tests of Significance for Effect GRA

First Data Set

Test Name	Value	Approx. F	df	p
Pillai's	1.73333	39.00000	4, 24	.000
Hotelling's	34.00000	85.00000	4, 20	.000
Wilks'	.00741	58.40423	4, 22	.000
Roy's	.96850			

Table 4

Eigenvalues for Effect GRA

First Data Set

Root No.	Eigenvalue
1	30.74773
2	3.25227

Table 5

Multivariate Tests of Significance for Effect GRA

Second Data Set

Test Name	Value	Approx. F	df	p
Pillai's	.63351	2.78161	4, 24	.050
Hotelling's	1.72713	4.31783	4, 20	.011
Wilks'	.36663	3.58336	4, 22	.021
Roy's	.63328			

Table 6
Eigenvalues for Effect GRA
Second Data Set

Root No.	Eigenvalue
1	1.72691
2	.00022

Table 7
Multivariate Tests of Significance for Effect TRE
Second Data Set

Test Name	Value	Approx. F	df	p
Pillai's	.81702	24.55814	2, 11	.000
Hotelling's	4.46512	24.55814	2, 11	.000
Wilks'	.18298	24.55814	2, 11	.000
Roy's	.81702			

Table 8
Eigenvalues for Effect TRE, Second Data Set

Root No.	Eigenvalue
1	4.46512

Table 9
Discriminant Functions on Groups Defined by GRA
After
Function Wilks' Lambda Chi-squared df Significance

0	0.0074074	71.126	4	0.0000
1	0.2351683	20.988	1	0.0000

Table 10

Structure Matrix

	Function 1	Function 2
DEPA	-0.41554	0.90958
DEPB	0.57995	0.81466

Table 11

Group Centroids

Group	Function 1	Function 2
1	-5.83183	1.35022
2	-0.67950	-2.31768
3	6.51133	0.96745

Table 12

Group Means

GRA	DEPA	DEPB
1	8	4
2	4	4
3	4	9